

Introducing theinfo.org

A lot of the work I've been doing on Open Library for the past few months has to do with handling large quantities of data. Either I'm writing crawlers to download them from various public web sites, or I'm meeting with librarians to persuade them to give me copies, or I'm evaluating algorithms for processing them, or building tools for viewing it all.

And while I've been doing this for information about books, I've noticed my friends doing similar things in other fields. Reporters try to get large data sets to write stories. Programmers get large data sets to add features to their sites. Friends are trying to make available data about the inner workings of the government.

And while each community has ways of talking to each other — reporters talking to other reporters, RDF people talking to other RDF people, library hackers talking to other library hackers — there's no community that cuts across these topical lines. And that's too bad, because there's a lot there we could share, from tips on how not to get caught when crawling to tools to make it easier to build big charts and maps.

So that's why I've started a new community site for people who work with large data sets. It's called theinfo.org and I'd really appreciate it if you joined the mailing lists and spread the word.

<http://theinfo.org/>

You should follow me on twitter [here](#).

January 15, 2008